SciDataCon 2016

Conference paper

USGS Data Management Website: Helping our scientists

Michelle Chang, U.S. Geological Survey

Abstract

In 2013, a series of White House orders required that federal data be made open and machine readable to the public. In response, the U.S. Geological Survey (USGS) developed a Public Access Plan and internal policies to meet these open data requirements. Launched in 2012, the USGS Data Management Website provides non-prescriptive guidance, best practices, and tools for good data management which are essential to making data more accessible, discoverable, and re-usable. Structured around the USGS Science Data Lifecycle Model, the website provides guidance for each stage of the data lifecycle from planning for data to preserving and sharing data at the end of a science project. For example, the website includes how to write a data management plan, develop data standards, create robust metadata, use persistent identifiers, preserve digital data and physical samples, and publish and share data. The USGS Data Management Website, available at www.usgs.gov/datamanagement, receives thousands of site visitors every month, some of which have led to successful partnerships with other government agencies to improve data management and make data more open. For questions and comments, please contact GS_Data_Management@usgs.gov.

Background

In 2013, President Obama signed the Executive Order “Making Open and Machine Readable the New Default for Government Information” to make data collected by the government more open and accessible (Obama 2013). This along with White House memoranda released by the Office of Science and Technology Policy called “Increasing Access to the Results of Federally Funded Scientific Research” (Holdren et al. 2013) and the Office of Management and Budget called “Open Data Policy – Managing Information as an Asset” (Burwell et al. 2013) support transparency by addressing the dissemination and interoperability of data and information products produced by federal agencies.

In response, the U.S. Geological Survey (USGS) published the plan “Public Access to Results of Federally Funded Research at the U.S. Geological Survey”, which discusses the USGS’s unified approach to make federally funded data more available (USGS 2016). Moreover, USGS has enacted a series of Instructional Memoranda (IM) which provide specific data management and data publishing policies (USGS 2015) for its staff to successfully meet the new open data requirements.

Bringing Good Data Management Practices to Our Science

Data management directly facilitates open data. While data management is an essential component of any successful science project, data management can help to ensure that the

data will be preserved and shared beyond the life of that project. Data that are well documented and easily accessed may be integrated more readily with other data, which reduces redundant work and adds value. Furthermore, well-documented data can be shared with and re-used by other scientists now, and in the future, and can be used to support the integrity and transparency of the science.

The USGS Data Management Website plays an important role in helping to achieve the objectives of the open data effort by focusing on good data management practices. Launched in 2012 by the USGS Community for Data Integration (CDI) Data Management Working Group, the website provides nonprescriptive data management guidance, best practices, and tools. The website was developed on the premise that well-managed, accessible data is crucial in order to perform data integration activities and produce analysis to address challenging science questions in this era of big data. The CDI identified a need for USGS scientists and their partners to have access to standard procedures, guidelines, and best practices in data management available in one place.

The website is organized around the USGS Science Data Lifecycle Model (Figure 1), a foundation and framework for managing data - from planning through preservation and sharing, and including documentation, quality assurance, and securing of data (Faundeen et al. 2013). The Data Lifecycle process is complementary to the research process; performing data management activities during science projects enables scientists to align data management requirements at the appropriate stages in their individual, respective projects. Users can learn about elements of the Data Lifecycle, to uncover more detailed topics, and easily locate best practices, tools, handouts, and recommended readings. Tying the website to the Data Lifecycle is a way to help users understand how each element fits in the scheme of data management, and how it may be applied.

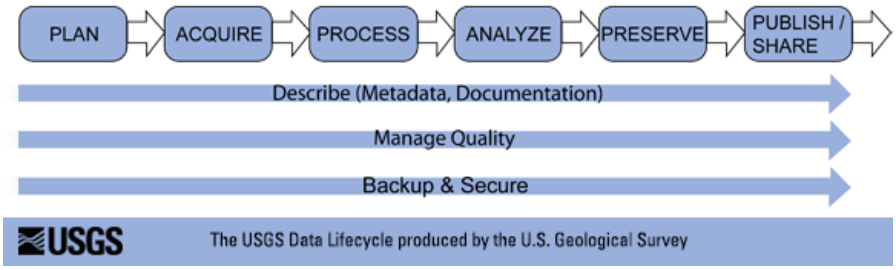

Figure 1: Depiction of the USGS Science Data Lifecycle Model used on the USGS Data Management Website. This graphic was modified from the USGS Science Data Lifecycle Model report (Faundeen et al. 2013).

The USGS IM require certain provisions such as a data management plan for science projects, robust metadata that describe data, a persistent identifier that links to the location of data, and a formal review process for the release of science data to the public (USGS 2015). All of these elements are good data management practices that provide the foundation for making data more accessible, discoverable, and re-usable.

The USGS Data Management Website supports the USGS IM by providing guidance on how to:

- Create a data management plan
- Develop data standards, data formats, and data templates
- Create compliant and robust metadata records
- Assign persistent identifiers (e.g. Digital Object Identifiers) to datasets for citation and better tracking
- Preserve and archive digital data and physical samples

- Publish and share data to promote discoverability, access, and reuse
- Learn the basics of data management through training modules
- Meet the new USGS data release requirements for publishing data to the public

Outcomes and Benefits

The website has received recognition for providing quality data management resources to USGS and the public. The USGS awarded the site the 2013 USGS Eugene M. Shoemaker Award for excellence in communications and the site has also received numerous accolades from other government agencies for its content and resources. These have led to collaborative efforts to improve data management knowledge across agencies. For example, USGS and the U.S. Air Force Research Laboratory established a Memorandum of Understanding to help create an online data management resource directed at the aerospace materials research and engineering community. In addition, website visits have quadrupled over the past four years with thousands of visitors per month.

USGS Earth science data is an invaluable asset, critical to informing decision makers on how to address large and complex societal issues. Through partnerships, user feedback, and contributions from scientists and data managers, the USGS Data Management Website will continue to grow as a dynamic resource for sharing best practices and leveraging solutions that help make data more open. We must continue to help keep our science relevant by making data easily accessed, well documented, and readily reusable. For questions and comments, please contact GS_Data_Management@usgs.gov.

Acknowledgements

The author would like to thank Viv Hutchison, Heather Henkel, and Lisa Zolly for their data management expertise and support. The author would also like to thank the Community for Data Integration (CDI) and the CDI Data Management Working Group members for their ongoing contributions to the USGS Data Management Website.

Competing Interests

The author declares that they have no competing interests.

References

Burwell, S M, Mancini, D J, Park, T, and VanRoekel, S 2013 Open Data Policy—Managing Information as an Asset: Executive Office of the President, Office of Management and Budget, Memorandum M-13-13 (May 9, 2013), 12 p. Available at http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m- 13-13.pdf [Last accessed 9 May 2016].

Faundeen, J L, Burley, T E, Carlino, J A, Govoni, D L, Henkel, H S, Holl, S L, Hutchison, V B, Martín, E, Montgomery, E T, Ladino, C C, Tessler, S, and Zolly, L S 2013 The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p. DOI: http://dx.doi.org/10.3133/ofr20131265.

Holdren, J P 2013 Memorandum for the heads of executive departments and agencies—Increasing Access to the Results of Federally Funded Scientific Research: Executive Office of the President, Office of Science and Technology Policy (February 22, 2013), 6 p. Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf [Last accessed 9 May 2016].

Obama, B 2013 Executive Order No. 13642, Making Open and Machine Readable the New Default for Government Information: Federal Register Vol. 48, No. 93, pg. 28111 (May 9, 2013), Available at https://www.federalregister.gov/articles/2013/05/14/2013-11533/making-open-and-machine-readable-the-new-default-for-government-information [Last accessed 9 May 2016].

U.S. Geological Survey 2015 Fundamental Science Practices Policies: Office of Science Quality and Integrity (OSQI) Instructional Memoranda (IM) - IM OSQI 2015-01, IM OSQI 2015-02, IM OSQI 2015-03, IM OSQI 2015-04. Available at http://www2.usgs.gov/fsp/policies.asp [Last accessed 9 May 2016].

U.S. Geological Survey 2016 Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data. Available at http://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf [Last accessed 9 May 2016].